

Transformadas direccionales basadas en grafos para codificación de vídeo: hacia un nuevo paradigma en codificación.

Eduardo Martínez Enríquez

Junio 2018

Capítulo 1: Introducción

Esta línea de investigación comenzó durante mi estancia predoctoral en la universidad *University of Southern California* (Los Ángeles) bajo la supervisión del profesor Antonio Ortega.

El objetivo principal es el diseño y desarrollo de nuevos codificadores de vídeo basados en transformadas direccionales construidas mediante grafos. La idea es novedosa y los resultados de codificación obtenidos hasta el momento son prometedores [1-5]. Además, se puede extender fácilmente a otros ámbitos de aplicación (como por ejemplo, codificación de audio multicanal o “denoising” de señales N-dimensionales). No obstante, dada la novedad de la propuesta, aún hay mucho margen de mejora y problemas que resolver para obtener un codificador que sea competitivo en entornos más realistas, especialmente en cuanto a complejidad computacional y optimización se refiere.

El capítulo siguiente describirá con detalle los objetivos y resultados esperados de esta línea de investigación así como la estructura del proyecto de investigación.

Capítulo 2: Línea de investigación

En el presente capítulo se introducirá la línea de investigación “Transformadas direccionales basadas en grafos para codificación de vídeo: hacia un nuevo paradigma en codificación”, describiendo la motivación y objetivos, así como desglosando las diferentes partes en las que se puede descomponer el proyecto.

Introducción

Las transformadas clásicas como la transformada discreta del coseno o la transformada Wavelet discreta pierden eficacia cuando se aplican a señales D-dimensionales con grandes discontinuidades (por ejemplo, los bordes de una imagen), dando lugar a coeficientes de valor alto en estas posiciones.

En tales casos, las transformadas direccionales, que son capaces de adaptar sus funciones base a la estructura de señal subyacente evitando filtrar a través de grandes discontinuidades, pueden dar lugar a un mejor rendimiento, compactando la señal en un número menor de coeficientes de valor alto.

El diseño de transformadas direccionales ha sido un campo de investigación activo en las últimas dos décadas. Ejemplos de este tipo de transformadas son las Curvelets [6], Contourlets [7], Bandelets [8], Directionlets [9] o DCTs direccionales [10, 11].

En trabajos previos diseñamos un codificador de vídeo basado en transformadas direccionales construidas a partir de la representación de la señal de vídeo en grafos. En dicha representación, cada nodo del grafo se asocia a un píxel y los enlaces entre los nodos capturan similitud entre los valores de luminancia. Esto se puede realizar, por ejemplo, enlazando píxeles relacionados por un proceso de estimación de movimiento o enlazando píxeles vecinos (pertenecientes al mismo cuadro) que no correspondan al borde de una imagen (que pertenezcan al mismo objeto). La figura 1 muestra este proceso.

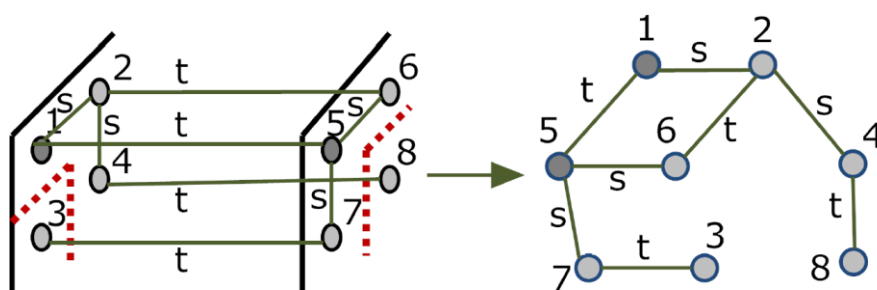


Figura 1. Construcción del grafo a partir de la señal de vídeo. s=enlaces espaciales; t=enlaces temporales.

Una vez se tiene la representación de la señal de vídeo en un grafo, se lleva a cabo una transformada “lifting” sobre dicho grafo [12, 13].

Básicamente, para realizar la transformada, el grafo construido se divide en dos subgrafos disjuntos (es decir, se asigna una etiqueta a cada nodo del grafo, de manera que sean nodos de actualización (U) o de predicción (P)). A continuación se definen los filtros de actualización (**u**) y predicción (**p**). Luego, en la etapa de predicción de la transformada, se predicen los nodos P a partir de los nodos U utilizando los filtros de predicción **p**, obteniendo versiones submuestreadas paso alto (coeficientes de detalle) de la señal original; finalmente, en la etapa de actualización las muestras U se actualizan a partir de coeficientes de detalle P utilizando los filtros **u**, dando lugar a una versión paso bajo submuestreada de la señal (coeficientes de suavizado). Si los coeficientes de detalle son cercanos a cero, la información principal de la señal se almacena en los coeficientes de suavizado (paso bajo), obteniendo así una representación más compacta. Aplicando este proceso iterativamente en los coeficientes paso bajo se obtiene un análisis multirresolución (MRA) [14] de la señal original. La figura 2 resume la transformada “lifting” sobre grafos.

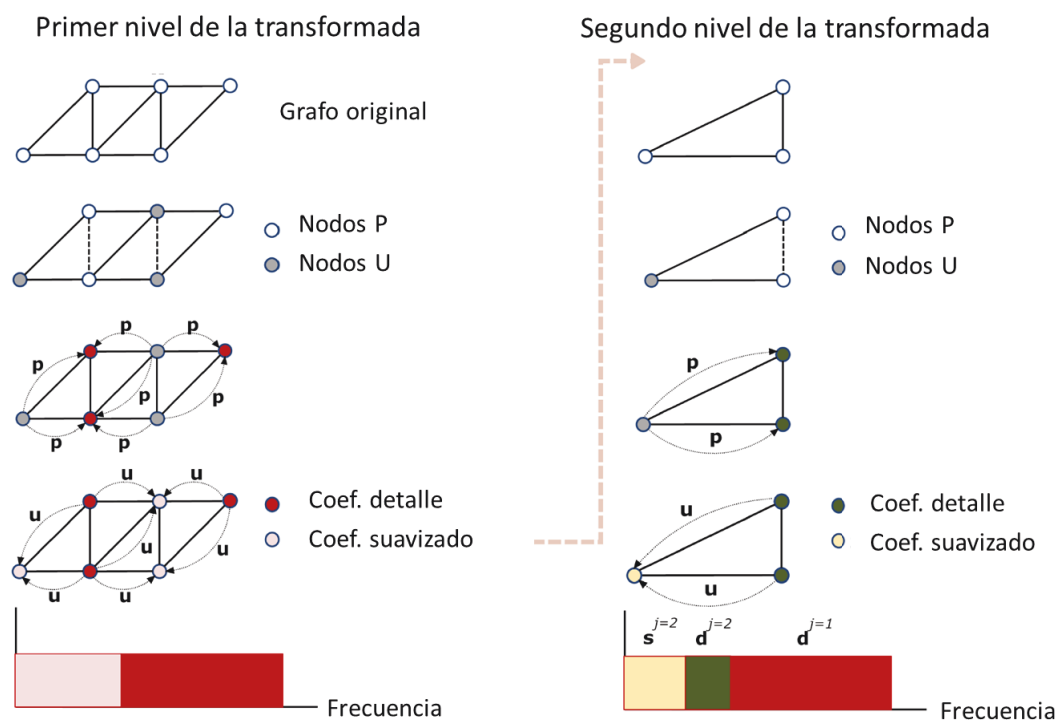


Figura 2. Transformada “lifting” sobre grafos.

Es importante denotar que las operaciones de filtrado se realizan haciendo uso de los nodos enlazados en el grafo y por tanto siguiendo las direcciones de alta similitud, consiguiendo así la direccionalidad de la transformada.

La novedad del método propuesto se resume en los siguientes puntos:

- Las principales estructuras de codificación de video basadas en “lifting” son denominadas “ $t + 2D$ ” y “ $2D + t$ ”. En la primera se aplica primero la transformada “lifting” en el dominio temporal, filtrando a lo largo de un conjunto de trayectorias de movimiento descritas por un modelo de movimiento específico (enfoque conocido como filtrado temporal compensado por movimiento, MCTF). A continuación, se lleva a cabo una transformada wavelet¹ bidimensional en el dominio espacial [15]. En la segunda, cada cuadro es primero transformado en el dominio espacial, seguido de MCTF. Ejemplos representativos de implementaciones MCTF son [16-18]. **Estos enfoques se pueden describir como separables porque el filtrado espacial y temporal se aplica en pasos separados.** El trabajo que propusimos en [1-4], gracias al uso de grafos, da lugar a transformadas no separables que explotan la redundancia espacial y temporal simultáneamente.
- En los trabajos presentados en el estado del arte, para realizar las etapas de predicción y actualización del esquema “lifting”, la secuencia a codificar se divide en subsecuencias U (cuadros pares) y P (cuadros impares), y para cada nivel de la transformada, la subsecuencia P es predicha a partir de la subsecuencia U dando lugar a la subbanda paso alto, y la subsecuencia de U se actualiza utilizando una versión filtrada de la P, obteniendo así la secuencia paso bajo. En los casos en que el modelo utilizado para describir el movimiento no pueda capturar con precisión el movimiento “real” de la escena, este tipo de división en cuadros pares e impares conducirá a enlazar píxeles de U y P con valores de luminancia muy diferentes. Debido a esto, los cuadros P serán mal predichos desde imágenes U, dando lugar a coeficientes de alta energía en la

¹ Nótese que la transformada “lifting” es una manera directa y sencilla de construir transformadas Wavelet.

secuencia paso alto, y por lo tanto a baja compactación de la energía. Además, cuando se utiliza MCTF, surgen problemas debidos a oclusiones y áreas descubiertas (es decir, píxeles que se filtran varias veces o que no se filtran en ningún momento). Con el esquema propuesto los píxeles se enlazan con un número arbitrario de vecinos espaciales y temporales simultáneamente, resolviendo estos problemas de manera sencilla y natural.

El método de codificación propuesto ha despertado interés en el campo de codificación de vídeo, posiblemente, gracias a su novedad. Así, trabajos en los que se presentó esta nueva metodología por primera vez recibieron importantes distinciones como el *Best paper award* en ICIP 2011 o el *Top 10 % paper recognition* en ICIP 2013. Además, en [1] se presenta un codificador completo que obtiene resultados prometedores en cuanto a eficiencia de codificación se refiere. Es dicha novedad y los resultados prometedores lo que me motivan a proponer esta línea de investigación.

Objetivo del proyecto de investigación

Nuestra nueva manera de codificar vídeo ha mostrado mejoras respecto a codificadores MCTF e incluso respecto a codificadores basados en H.264 configurados en las mismas condiciones de trabajo [1]. La figura 3 muestra un ejemplo de la calidad subjetiva obtenida codificando con el método propuesto (Fig3.b), H.264 simplificado (Fig3.c), y con un enfoque MCTF (Fig3.d)



(a)



(b)



(c)



(d)

Figura 3. Calidad subjetiva obtenida codificando un cuadro (Fig3.a) con el método propuesto (Fig3.b), H.264 simplificado (Fig3.c), y con un enfoque MCTF (Fig3.d).

No obstante, las condiciones de trabajo consideradas en este caso para H.264 serían una simplificación de las posibilidades del estándar (por ej., estimación de movimiento en tamaños 16x16 únicamente y con precisión de píxel), por lo que no se obtiene un codificador realista. El codificador propuesto hasta el momento obtiene peores resultados que H.264 funcionando al máximo de sus prestaciones [1].

El objetivo del proyecto de investigación es **generar una nueva manera de codificar vídeo que sea novedosa y competitiva con el estado del arte y los estándares de codificación actuales**. Para ello, se deberán cumplir las siguientes características:

- **Novedad**
- **Eficiencia de codificación**
- **Eficiencia en cuanto a coste computacional se refiere.**

Para mejorar la **eficiencia de codificación** se trabajará en el desarrollo de nuevas herramientas (como incluir la estimación de movimiento con precisión subpíxel) en el

nuevo codificador. Sin embargo, la **novedad** del método lleva implícita la necesidad de desarrollar herramientas totalmente nuevas (distintas a las propuestas en el estado del arte) en cada parte del codificador, como el desarrollo de un nuevo codificador entrópico que haga uso de la información del grafo, o de una manera de optimizar los parámetros del codificador (incluyendo la asignación de nodos U/P y el diseño de los filtros **u** y **p**) desde un punto de vista tasa-distorsión. Por ejemplo, en [3] se propuso una nueva forma de reordenar los coeficientes (intra e inter-subbanda) utilizando la información del grafo que condujo a una mejora considerable en eficiencia respecto a versiones anteriores del codificador. Finalmente, para que el codificador sea realista habrá que cuidar en todo momento el **coste computacional** de los algoritmos propuestos, especialmente al trabajar con grafos enormes de tamaño Número de píxeles/imagen x Número de imágenes consideradas en la secuencia de vídeo.

Estructura del proyecto de investigación

Para alcanzar los objetivos descritos previamente, se propone una estructura para el proyecto como la que se muestra en la figura 4.

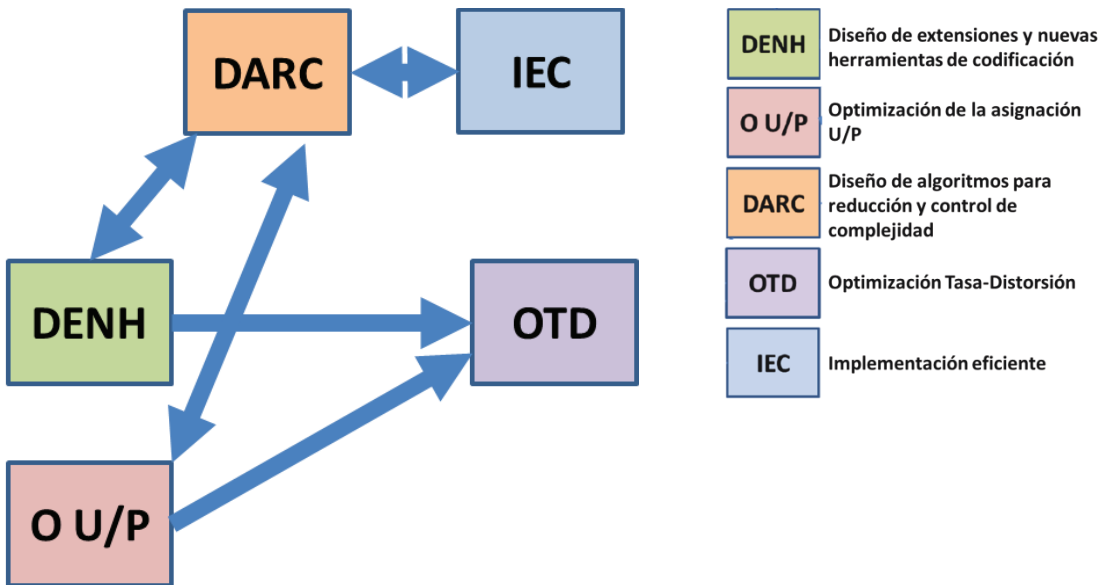


Figura 4. Estructura propuesta para el proyecto de investigación.

El primer bloque es el **diseño y desarrollo de extensiones y nuevas herramientas de codificación (DENH)**, cuyo objetivo principal sería aumentar la eficiencia de codificación proponiendo novedosas herramientas.

El segundo bloque será la optimización de la **asignación U/P (OUP)**. Dicha optimización será una de las más importantes para conseguir mejorar la eficiencia de codificación y reducir la complejidad. Por lo tanto, se intentará desarrollar nueva teoría y algoritmos prácticos que nos ayuden a comprender la forma óptima de asignar una etiqueta U o P a cada nodo del grafo. Nótese que este bloque se puede trabajar en paralelo con el descrito anteriormente (DENH).

Una vez se tienen las distintas opciones de codificación posibles, se trabajará en la **optimización tasa-distorsión (OTD)** para obtener los parámetros que minimizan la distorsión dada una restricción de tasa.

Por otro lado, tendremos el bloque de **diseño de algoritmos para reducción de complejidad (DARC)**, que pretenderá reducir la complejidad de los algoritmos de manera que se obtenga un codificador útil. Este bloque irá muy ligado al de **implementación eficiente del codificador (IEC)**, donde se contemplarán aspectos más relacionados con implementación y posible desarrollo hardware de un prototipo del codificador. Este bloque no será tratado en el presente documento.

Es importante resaltar que muchas tareas pueden desarrollarse en paralelo, y que siempre habrá que tener en cuenta el coste computacional de las soluciones propuestas (especialmente en la asignación U/P).

A continuación se describen las posibles etapas del proyecto de manera pormenorizada, indicando posibles subtareas y objetivos parciales de cada una de ellas.

1. Diseño de extensiones y nuevas herramientas de codificación (DENH).

Como se ha comentado anteriormente, el punto de partida del proyecto sería el codificador propuesto en publicaciones previas [1-4]. En los experimentos de [1] se muestra cómo dicho codificador es capaz de obtener una eficiencia similar o superior a un codificador que emplee las mismas herramientas de codificación en H.264 (por ejemplo, no usar subpixel o hacer la estimación de movimiento en bloques de tamaño 16x16). Sin embargo, estas condiciones no son realistas y el objetivo será comparar el codificador propuesto con la versión más completa posible de los últimos estándares de codificación (tamaño de bloque variable, múltiples referencias, codificación aritmética sofisticada, ...).

Los últimos estándares de codificación de vídeo incluyen múltiples herramientas para mejorar la eficiencia de compresión, como emplear varios cuadros de referencia; diferentes modos de predicción y tamaños de partición; estimación de movimiento con precisión de subpixel o incluso mayor; o codificadores entrópicos adaptativos al contexto (CAVLC o CABAC) por poner algunos ejemplos. Dichas herramientas mejoran significativamente la eficiencia del codificador. Además, generalmente se realiza un proceso de optimización tasa-distorsión (aunque éste no está estandarizado) que permite al codificador elegir la mejor opción de codificación entre las diferentes combinaciones de las herramientas descritas anteriormente [19, 20].

Por el momento, la cantidad de herramientas diseñadas e implementadas en el codificador propuesto es limitada. Sin embargo, gracias a su versatilidad, sería posible combinar nuestro enfoque con muchas de las herramientas mencionadas anteriormente u otras que diseñáramos específicas para trabajar con señales definidas en grafos.

En transformadas “lifting” sobre grafos un determinado nodo se puede enlazar con cualquier subconjunto de nodos sin restricciones manteniendo las propiedades de reconstrucción perfecta (transformada completamente invertible) y muestreo crítico (mismo número de coeficientes que de píxeles de la señal original) de la transformada.

Por lo tanto, usar **bipredicción**, **aumentar el número de referencias** o incorporar los **tamaños de partición variables** podría hacerse de una manera aparentemente sencilla simplemente creando los enlaces correspondientes entre nodos y enviando la información adicional necesaria al decodificador (de manera que sea capaz de reconstruir el mismo grafo para realizar la transformada inversa y así decodificar la señal). La predicción ponderada también puede ser incorporada gestionando los pesos de los enlaces entre nodos. Además, como se muestra en [21, 22], la **estimación de movimiento con precisión de subpixel** puede aplicarse en implementaciones de MCTF sin perder la invertibilidad de la transformada. Habría que investigar de qué manera y bajo qué condiciones esta teoría se puede extender a nuestro sistema² para que la estimación de movimiento subpixel sea invertible usando muestras interpoladas (en los dominios espaciales o temporales) de los nodos U.

La figura 5 muestra de manera gráfica e intuitiva la forma de incluir varias de estas herramientas en el codificador propuesto.

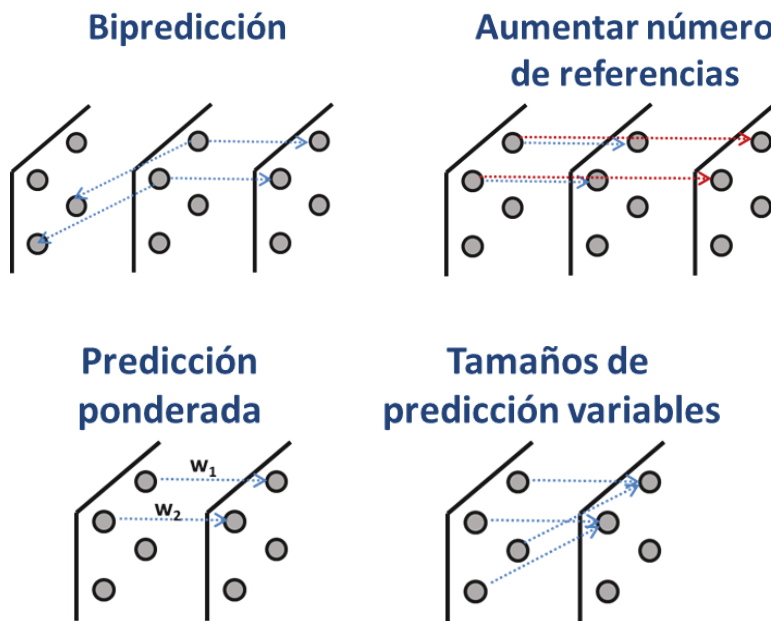


Figura 5. Bipredicción, predicción ponderada, tamaño de predicción variable y múltiples referencias con el método de codificación propuesto.

² Recuérdese que nuestro sistema trabaja con grafos arbitrarios donde cada nodo puede tener un número cualquiera de vecinos U y P.

Otra vía interesante de investigación sería mejorar el **codificador entrópico**. Mientras que el propuesto en [1] es similar a CAVLC, se podría mejorar explotando correlaciones entre subbandas [23, 24] o haciendo dicho codificador adaptativo al contexto como en H.264/AVC, donde se elige entre diferentes “look-up tables” en función del contexto (por ejemplo, número de coeficientes distintos de cero en bloques vecinos o magnitudes de nivel recientemente codificadas). Finalmente, se podría profundizar en la manera de **reordenar los coeficientes** aprovechando la información del grafo transmitida al decodificador como se hizo en [3].

2. Optimización de la asignación U/P (OUP).

Nuestra propuesta obtiene transformadas direccionales mediante la aplicación de transformadas “lifting” sobre grafos. Estas son muy útiles para obtener de manera muy simple una representación eficiente invertible y críticamente muestreada de una señal dada en un grafo. Sin embargo, su eficiencia depende fuertemente de diferentes problemas de optimización cuyas soluciones no son triviales cuando se consideran grafos arbitrarios. Como se ha comentado anteriormente, el proceso de asignación U/P, que consiste en asignar una etiqueta a cada nodo del grafo, es uno de los procesos más relevantes de la transformada. En [1] se demuestra experimentalmente la importancia del proceso de asignación U/P, evaluando la capacidad de compactar la energía del codificador cuando se usan distintas estrategias de asignación U/P.

En [5] definimos específicamente el problema de optimización e introducimos, por primera vez, teoría en la optimización de la asignación U/P en transformadas “lifting” sobre grafos, proponiendo una metodología general de asignación U/P que minimiza el valor esperado del error de predicción. Específicamente, se demostró que dado un grafo arbitrario y un modelo de señal (consideramos “moving average models”) la asignación óptima depende de la correlación entre los nodos del grafo, y no es la que maximiza el corte del grafo como se propone en trabajos previos [1]. Esta conclusión demostrada teóricamente se validó experimentalmente evaluando la energía media en los coeficientes de detalle cuando se utiliza la nueva metodología de asignación U/P y

otras metodologías propuestas en el estado del arte (como la propuesta en [1]). Dichos experimentos se realizaron en grafos generados a partir de señales aleatorias sintéticas y en grafos generados a partir de datos reales provenientes de señales de imagen y video. El objetivo de este bloque sería aplicar este avance teórico al codificador de vídeo, abordando algunas consideraciones práctico/teóricas aún no resueltas, como:

- Adaptación de dicha asignación U/P al codificador de vídeo [1], para lo cual habrá que reducir el coste computacional del método propuesto mediante la aplicación de algoritmos explorados en [3, 4] y una nueva serie de soluciones.
- Investigar la asignación óptima empleando modelos de señal más realistas a la hora de describir la señal de imagen y vídeo (como "*Gaussian Markov Random Fields*").
- Hacer una optimización multinivel que no considere independencia entre los distintos niveles de la transformada (ya que el resultado en uno de los niveles influirá en los siguientes).

3. Diseño de algoritmos para reducción y control de complejidad (DARC)

A día de hoy el uso de dispositivos multimedia portátiles está totalmente extendido. Por lo tanto, los sistemas de procesamiento de señal (y en nuestro caso concreto de codificación de vídeo) deben diseñarse de manera que funcionen correctamente en una gran variedad de plataformas, cada una dotada de una capacidad de computación y recursos de memoria específicos. Por lo tanto, la concepción de algoritmos capaces de adaptar su complejidad computacional (obviamente a cambio de rendimiento, memoria, retardo, etc.) a la requerida por dispositivos específicos se convierte en un desafío importante.

En el entorno de H.264/AVC y HEVC la complejidad es muy alta y viene dada por la gran variedad de herramientas de codificación disponibles y la necesidad de evaluar

distintos parámetros de cada herramienta con el fin de seleccionar la mejor combinación desde el punto de vista tasa-distorsión. Por ello, existen numerosos métodos en el estado del arte que tratan de reducir la complejidad de codificadores bajo estos estándares [25-32]. La mayor complejidad del método propuesto viene dada por el proceso de asignación U/P.

A día de hoy, la implementación *Matlab* del codificador es aproximadamente el doble de compleja (cuando se evalúa tiempo de codificación) que la de un sistema MCTF comparable también implementado en *Matlab*. Los objetivos preliminares para este bloque serán los siguientes:

- A pesar de que para realizar la asignación U/P se reduce en gran medida la complejidad del codificador empleando el enfoque distribuido presentado en [3] y el enfoque “sub-grafo” de [4] (donde se trabaja por subgrafos generados a partir del grafo general), aún hay un gran margen de mejora en cuanto a reducción de complejidad.

Posibles soluciones serían (i) trabajar en pequeños subgrafos de manera que el hecho de crear particiones del grafo original afecte lo mínimo posible a la eficiencia de codificación; (ii) reducir aún más la complejidad y ser capaces de controlarla limitando el tamaño de los subgrafos; (iii) usar el procesamiento paralelo (realizar, por ejemplo, la asignación de U/P para los diferentes bloques de manera paralela e intercambiar la información necesaria final en las fronteras de dichos bloques para conseguir una solución global coherente).

- El número promedio de vecinos temporales por nodo son 2 para cada secuencia, pero la desviación estándar suele ser más alta para secuencias con gran movimiento. Además, el número de vecinos espaciales por nodo varía mucho entre las secuencias, siendo más bajo en las secuencias con alta textura o gran cantidad de contornos. Por lo tanto habrá que tener en cuenta el hecho de que el tiempo de codificación es muy variable a la hora de trabajar con distintas secuencias e intentar minimizar este efecto.

- Extender las técnicas de reducción/control de complejidad empleadas en estándares de codificación para, por ejemplo, seleccionar el tamaño de partición óptima de manera prematura.

4. Optimización Tasa-Distorsión (OTD).

En una aplicación de codificación típica, el codificador realiza una optimización de tasa-distorsión (RDO) para encontrar los parámetros de codificación que minimizan la distorsión bajo una restricción de tasa [20, 33-35]. Igualmente, para el esquema de codificación propuesto es necesario optimizar los parámetros de esta manera. Especialmente, el número de nodos P en el grafo (que tiene una gran influencia en la eficiencia de codificación) debe elegirse resolviendo el RDO.

En [36] primero se formula el problema original con restricciones de RDO convirtiéndolo en problema sin restricciones. Luego, se proponen modelos teóricos de tasa (R) y distorsión (D) que dependen de la asignación U/P (que implícitamente determina el número de nodos P), y el escalón de cuantificación Δ de los coeficientes de suavizado para el primer nivel de la transformada. Aunque se necesitaron algunas simplificaciones para construir los modelos, estos dan una intuición muy útil sobre cómo aplicar RDO para optimizar el número de nodos P en el grafo. A continuación, usamos los modelos propuestos para R y D para obtener analíticamente el parámetro λ que equilibra el peso de los términos tasa y distorsión en el problema de RDO sin restricciones. De esta manera, obtuvimos una fórmula que relaciona λ y Δ , quedando un único parámetro libre en el proceso de optimización. Finalmente, propusimos un algoritmo para realizar dicho proceso de optimización.

Por lo tanto, la investigación a llevar a cabo en esta línea se podría describir de la siguiente manera:

- El trabajo realizado en [5] busca la partición U/P óptima (que minimiza la energía de los coeficientes de detalle) para un número de nodos P dado. Sin embargo, en una aplicación práctica, el número de nodos P no se conoce α

priori y será de crucial interés seleccionarlo en función de la aplicación o, tratándose de codificación de vídeo, en función de las características del vídeo a codificar. Para ello tendremos que optimizarlo utilizando nuestra formulación de RDO.

- El proceso mostrado en [36] asume ciertas simplificaciones. Será necesario extender dicho trabajo eliminando algunas de estas simplificaciones para obtener un proceso más realista y por lo tanto un codificador más eficiente.
- Será necesario realizar la optimización evitando que se produzcan grandes discontinuidades en los parámetros seleccionados -y con ello efectos subjetivos no deseados- o en tal caso, proponer soluciones para paliar dichos efectos como las presentadas en trabajos previos [37, 38].

Referencias

- [1] E. Martínez-Enríquez, J. Cid-Sueiro, F. Díaz-De-Maria, and A. Ortega, "Directional transforms for video coding based on lifting on graphs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, pp. 933-946, November 2016.
- [2] E. Martínez-Enríquez and A. Ortega, "Lifting transforms on graphs for video coding," in *Data Compression Conference (DCC), 2011*, 2011, pp. 73-82.
- [3] E. Martínez-Enríquez, F. Díaz-de-María, and A. Ortega, "Video encoder based on lifting transforms on graphs," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 3509-3512.
- [4] E. Martínez Enríquez, F. Díaz-de-María, J. Cid Sueiro, and A. Ortega, "Filter optimization and complexity reduction for video coding using graph-based transforms," in *Image Processing (ICIP), 2013 IEEE International Conference on*, 2013, pp. 1948-1952.
- [5] E. Martínez-Enríquez, J. Cid-Sueiro, F. Díaz-de-María, and A. Ortega, "Optimized Update/Prediction Assignment for Lifting Transforms on Graphs," *IEEE Transactions on Signal Processing*, vol. 66, pp. 2098-2111, April 2018.
- [6] E. J. Candes and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*, ed: A. Cohen, C. Rabut, and L.L. Shumaker, Eds. Saint-Malo: Vanderbilt University Press, 1999.
- [7] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on image processing*, vol. 14, pp. 2091-2106, 2005.
- [8] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE transactions on image processing*, vol. 14, pp. 423-438, 2005.
- [9] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, and P. L. Dragotti, "Directionlets: anisotropic multidirectional representation with separable filtering," *IEEE Transactions on Image Processing*, vol. 15, pp. 1916-1933, 2006.
- [10] G. Fracastoro, S. M. Fosson, and E. Magli, "Steerable discrete cosine transform," *IEEE Transactions on Image Processing*, vol. 26, pp. 303-314, 2017.
- [11] B. Zeng and J. Fu, "Directional discrete cosine transforms—a new framework for image coding," *IEEE transactions on circuits and systems for video technology*, vol. 18, pp. 305-313, 2008.
- [12] S. K. Narang and A. Ortega, "Lifting based wavelet transforms on graphs," in *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, 2009, pp. 441-444.
- [13] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM journal on mathematical analysis*, vol. 29, pp. 511-546, 1998.
- [14] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, pp. 674-693, 1989.

- [15] N. Adami, A. Signoroni, and R. Leonardi, "State-of-the-art and trends in scalable video compression with wavelet-based approaches," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, p. 1238, 2007.
- [16] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE transactions on image processing*, vol. 12, pp. 1530-1542, 2003.
- [17] M. Flierl and B. Girod, "Video coding with motion-compensated lifted wavelet transforms," *Signal Processing: Image Communication*, vol. 19, pp. 561-575, 2004/08/01/ 2004.
- [18] G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans, "Motion compensation and scalability in lifting-based video coding," *Signal Processing: Image Communication*, vol. 19, pp. 577-600, 2004/08/01/ 2004.
- [19] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE signal processing magazine*, vol. 15, pp. 74-90, 1998.
- [20] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal processing magazine*, vol. 15, pp. 23-50, 1998.
- [21] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, 2001, pp. 1793-1796 vol.3.
- [22] S. Yea and W. A. Pearlman, "On scalable lossless video coding based on sub-pixel accurate MCTF," in *Electronic Imaging 2006*, 2006, p. 10.
- [23] B. J. Kim, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," *Proc. Data Compression Conf.*, pp. 251-260, 1997 1997.
- [24] J. M. Shapiro, "An embedded wavelet hierarchical image coder," in *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1992, pp. 657-660 vol.4.
- [25] A. Jimenez-Moreno, E. Martinez-Enriquez, and F. Diaz-de-Maria, "Mode decision-based algorithm for complexity control in H. 264/AVC," *IEEE Transactions on Multimedia*, vol. 15, pp. 1094-1109, 2013.
- [26] A. Jiménez-Moreno, E. Martínez-Enríquez, and F. Díaz-de-María, "Complexity Control Based on a Fast Coding Unit Decision Method in the HEVC Video Coding Standard," *IEEE Trans. Multimedia*, vol. 18, pp. 563-575, 2016.
- [27] E. Martínez-Enríquez, M. de-Frutos-López, J. C. Pujol-Alcolado, and F. Diaz-de-Maria, "A fast motion-cost based algorithm for H. 264/AVC inter mode decision," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 2007, pp. V-325-V-328.
- [28] E. Martínez-Enríquez and F. Diaz-de-Maria, "A hierarchical classification-based approach to inter mode decision in H. 264/AVC," in *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, 2009, pp. 221-224.
- [29] E. Martinez-Enriquez, A. Jimenez-Moreno, M. Angel-Pellon, and F. Diaz-de-Maria, "A two-level classification-based approach to inter mode decision in H. 264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, pp. 1719-1732, 2011.
- [30] E. Martinez-Enriquez, A. Jimenez-Moreno, and F. Diaz-de-Maria, "An adaptive algorithm for fast inter mode decision in the H. 264/AVC video coding

- standard," *IEEE Transactions on Consumer Electronics*, vol. 56, pp. 826-834, 2010.
- [31] E. Martínez-Enríquez, A. Jimenez-Moreno, and F. Diaz-de-Maria, "A novel fast inter mode decision in H. 264/AVC based on a regionalized hypothesis testing," in *Picture Coding Symposium, 2009. PCS 2009*, 2009, pp. 1-4.
 - [32] A. Jiménez-Moreno, E. Martínez-Enríquez, and F. Díaz-de-María, "Bayesian adaptive algorithm for fast coding unit decision in the High Efficiency Video Coding (HEVC) standard," *Signal Processing: Image Communication*, vol. 56, pp. 1-11, 2017/08/01/ 2017.
 - [33] J. L. González de Suso, A. Jiménez-Moreno, E. Martínez Enríquez, and F. Diaz-de-Maria, "Improved method to select the Lagrange multiplier for rate-distortion based motion estimation in video coding," *IEEE transactions on circuits and systems for video technology*, vol. 24, pp. 452-464, 2014.
 - [34] J. L. González-de-Suso, E. Martínez-Enríquez, and F. Díaz-de-María, "Adaptive Lagrange multiplier estimation algorithm in HEVC," *Signal Processing: Image Communication*, vol. 56, pp. 40-51, 2017/08/01/ 2017.
 - [35] J. L. Molinero, A. Jimenez, E. Martinez-Enriquez, and F. Diaz-de-Maria, "On the optimal Lagrangian parameter for motion estimation: A low-cost and effective method for improving video coding performance," in *Telecommunications (CONATEL), 2011 2nd National Conference on*, 2011, pp. 1-6.
 - [36] E. Martínez-Enríquez, "Lifting transforms on graphs and their application to video coding," Ph.D. dissertation, Universidad Carlos III de Madrid, 2013.
 - [37] A. Jiménez-Moreno, E. Martínez-Enríquez, and F. Díaz-de-María, "Standard compliant flicker reduction method with PSNR loss control," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 1729-1733.
 - [38] A. Jiménez-Moreno, E. Martínez-Enríquez, V. Kumar, and F. Díaz-de-María, "Standard-compliant low-pass temporal filter to reduce the perceived flicker artifact," *IEEE Transactions on Multimedia*, vol. 16, pp. 1863-1873, 2014.